CONSISTENCY OF GENERALIZED RANDOM DOT PRODUCT GRAPH WITH COVARIATES

by

Jipeng Zhang

A thesis submitted to The Johns Hopkins University
in conformity with the requirements for the degree of
Master of Science in Engineering

Baltimore, Maryland
December, 2019

Abstract

In this work we present Generalized Random Dot Product graph with Covariates model for network data with observed binary covariates. We introduce a spectral estimator for parameters of the model provided that the existence of edges in the graph are independently Bernoulli distributed and the latent positions of vertices are independent variables with some distribution $\mathcal{F}$. Theoretically, we prove that the estimator results are asymptotically equal to the true parameters up to some orthogonal transformations. Empirically, we utilize the Procrustes Procedures to find the exact orthogonal transformations. We investigate the algorithm to recover parameters for multiple binary covariates. We outline necessary related work and potential future work.

Thesis Committee

Primary Readers

Minh Tang (Primary Advisor)
Assistant Professor
Department of Statistics
North Carolina State University

Acknowledgments

Foremost, I would like to express the deep appreciation to my thesis advisor, Dr. Minh Tang, for the continuous support of my master's studies and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my master's research.

My sincere thanks also goes to Dr. Angelo Mele for offering me a research assistant position in his group and leading me working on graphical models for network analysis.

Also I am truly grateful to my academic advisor, Dr. Mauro Maggioni, for his warm encouragement and valuable suggestions throughout my master's studies.

Finally, I must express my very profound gratitude to my parents Fengmei Zhang and Tianlei Zhang, for giving birth to me at the first place and supporting me spiritually throughout my life.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Network analysis has broad applications in the field of economics, sociology, public health, and neuroscience. The idea of abstracting graphs with vertices and edges from networks could date back to 1736, when Leonard Euler proposed primary concepts of combinatorial graph theory to solve *Seven Bridges of Königsberg* problem. However, notations of random graph inference for modern research was not introduced until 1960, when Erdös and Rényi presented their graphical model (ER model) in which the existence of edges between vertices are independently Bernoulli distributed with a common probability. (Erdös et al., 1960) Regardless of the simplicity of the model, ER graphs showed astounding performance and considerable potential for analyzing network data. (Alon et al., 2008; Bollobás et al., 2007)

Nevertheless, the requirement of the same connection probability for any two vertices in ER model restricts its capacity. For example, the probabilities that two people connect on a social media should be different for distinct pairs of people. In 2002, Hoff et al. introduced latent position graph as a model for network data. (Hoff et al., 2002) In a latent position graph, each vertex is associated with a vector within the latent space $\mathbf{X}$ of $\mathbb{R}^d$, and there exists a kernel function $\kappa : \mathbf{X} \times \mathbf{X} \to [0, 1]$ giving the probability that two nodes are connected in a graph. Latent positions graph not only allows different likelihood of forming an edge, but also reveals vertex attributes by latent positions, which can be used to cluster communities of vertices. Based on Hoff et al.'s work, Young and Scheinerman proposed *random dot product graph* (RDPG), in which $\kappa(\mathbf{X}_i, \mathbf{X}_j) = \mathbf{X}_i^T \mathbf{X}_j$ with $\mathbf{X}_i, \mathbf{X}_j \in \mathbb{R}^d$. (Young et al., 2007) In RDPG, $\mathbf{P} = \mathbf{X}\mathbf{X}^T$ is necessarily positive semidefinite. To drop this requirement, Rubin-Delanchy, Tang, and Priebe raised *generalized random dot product graph* (GRDPG) in which kernel function $\kappa(\mathbf{X}_i, \mathbf{X}_j) = \mathbf{X}_i^T \mathbf{I}_{p,q} \mathbf{X}_j$. $\mathbf{I}_{p,q}$ is a diagonal matrix with p ones and q minus ones on its diagonal. (Rubin-Delanchy et al., 2017) In GRDPG, $\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^T$ does not have to be positive semidefinite.

For real-world networks, besides the adjacency matrix $\mathbf{A}$, usually some covariates could also be attainable. As an illustration, we can observe gender and race of all vertices for social networks. For two arbitrary vertices in a graph, it is reasonable that those covariates have an influence on the probability that there exists connection between these two vertices. Thus, a graphical model containing both unobserved latent positions and observed covariates information is desired to handle such real data. Mele et al. built stochastic blockmodels with one covariate and proposed a spectral algorithm to estimate parameters in the model. (Mele et al., 2019) However, Mele's algorithm has difficulties of being generalized from scenarios with one covariates to scenarios with multiple independent covariates, which is the main work of this thesis.

1.2 Problem Statement

Assume we have the adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ and covariates matrix $\mathbf{Z} \in \mathbb{R}^{n \times m}$, where $\mathbf{A}_{ij} = \mathbf{A}_{ji}$ when $i \neq j$ and $\mathbf{A}_{ii} = 0$, m covariates are binary and mutually independent. Then in the model of *generalized random dot product graph with covariates*, $\mathbf{A}$ is generated by the following process:

$$\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^T + \sum_{i=1}^{m} \beta_i \frac{\mathbf{1}\mathbf{1}^T + \mathbf{Z}^{(i)}\mathbf{Z}^{(i)T}}{2}$$

$$\mathbf{A}_{ij} | \mathbf{X}_i, \mathbf{X}_j \overset{ind}{\sim} Bernoulli(\mathbf{P}_{ij})$$

where $\mathbf{Z}^{(i)} \in \mathbb{R}^n$ denotes i^{th} column of $\mathbf{Z}$ and $\mathbf{Z}_j^{(i)} \in \{-1, 1\}$

Define $\boldsymbol{\beta} := [\beta_1, ..., \beta_m]^T \in \mathbb{R}^m$. Our goal is to estimate $\boldsymbol{\beta}$ from $\mathbf{A}$ and $\mathbf{Z}$, and prove consistency of our estimator.

1.3 Outline of the Thesis

The remaining of this thesis is organized as follows: Chapter 2 covers necessary related work and presents our methodology on random graph with covariates. Chapter 3 prove the consistency of GRDPG with covariates; more specifically, prove the estimated covariates parameters are close to ground truth up to orthogonal transformation. Chapter 4 describes the simulation procedures and presents some simulation results. Chapter 5 gives the conclusion of this thesis and discusses possible future research. Chapter 6 includes proofs of inequalities used in this thesis for completeness.

References

Erdös, P. and A. Rényi (1960). "On the Evolution of Random Graphs". In: *Proceeding of the Hungarian Academy of Sciences*, pp. 17–61.

Alon, N. and J. Spencer (2008). *The probabilistic Method*. John Wiley.

Bollobás, B., S. Janson, and O. Riordan (2007). "The phase transition in inhomogeneous random graphs". In: *Random Structures and Algorithms* 31, pp. 3–122.

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). "Latent space approaches to social network analysis". In: *Journal of the American Statistical Association* 97, pp. 1090–1098.

Young, S. and E. Scheinerman (2007). "Random dot product graph models for social networks". In: *Proceedings of the 5th international conference on algorithms and models for the web-graph*, pp. 138–149.

Rubin-Delanchy, P., C. E. Priebe, and M. Tang (2017). "A statistical interpretation of spectral embedding: the generalised random dot product graph". In: *arXiv:1709.05506*.

Mele, A., L. Hao, J. Cape, and C.E. Priebe (2019). "Spectral inference for large stochastic blockmodels with nodal covariates". In: *arXiv:1908.06438.*

Chapter 2

Related Work and Methodology

2.1 Stochastic Block Model

The *stochastic block model* (SBM) (Holland et al., 1983) is a widely used random graph model for heterogeneous social network data. In SBM, nodes of a network are devided into K communities. Let $\tau_i \in [1, ..., K]$ denote the community that vertex i belongs to. $\mathbf{B} \in [0,1]^{K \times K}$ is the characteristic matrix in which $\mathbf{B}_{ij}$ is the probability that a vertex from community i and a vertex from community j form a link in the network. Conditional upon the assigned communities, $\mathbf{A}_{ij} \overset{ind}{\sim} Bernoulli(\mathbf{B}_{\tau_i \tau_j})$. Moreover, the community assigned to an arbitrary vertex can be an independent and identically distributed random variable with some distribution on $\mathbb{R}$. In other words, $P(\tau_i = k) = \pi_k$ with $\sum_{k=1}^{K} \pi_k = 1$.

2.2 Generalized Random Dot Product Graph

Another graphical model for network data is *latent structure model* (LSM), which is more general than stochastic blockmodels due to the requirements on distributions. (Diaconis et al., 2007; Hoff et al., 2002; Smith et al., 2017) A *random dot product graph* (Young et al., 2007) (RDPG) is a latent positions graphical model. In RDPG, each node is characterized by a latent position of Euclidean space $\mathbb{R}^d$. Let $\mathbf{X}_i \in \mathbb{R}^d$ denote the latent position of vertex i. The probability that vertex i and vertex j form a link is $\mathbf{P}_{ij} = \mathbf{X}_i^T \mathbf{X}_j$. Conditional upon the latent positions, $\mathbf{A}_{ij} \overset{ind}{\sim} Bernoulli(\mathbf{X}_i^T \mathbf{X}_j)$. The matrix $\mathbf{P}$ is given by $\mathbf{P} = \mathbf{X}\mathbf{X}^T$, where $\mathbf{X} = [\mathbf{X}_1 | ... | \mathbf{X}_n]^T \in \mathbb{R}^{n \times d}$. The latent positions $\mathbf{X}_i$ may be an independent and identically distributed random variable with some distribution F on $\mathbb{R}^d$. Equivalently, $\mathbf{X}_1, ..., \mathbf{X}_n \overset{i.i.d}{\sim} F \in \mathbb{R}^d$.

Any SBM can be interpreted as a RDPG with the same latent position for all vertices in a given block. (Athreya et al., 2017) However, Note that in RDPG, $\mathbf{P}$ is necessarily positive semidefinite, which leads to one important assumption in RDPG model that $\mathbf{P}$ be positive semidefinite. To drop this assumption, Rubin-Delanchy, Tang, and Priebe proposed generalized random dot product graph (GRDPG). (Rubin-Delanchy et al., 2017) Without the restriction of positive semidefiniteness, GRDPG is more robust to real network data. Meanwhile, GRDPG holds many important useful attibutes of RDPG. (Athreya et al., 2017)

Definition 1. (GRDPG). Let $\mathbf{X}_i \in \mathbb{R}^d$ be the latent position of vertex i such that $\mathbf{X}_i^T \mathbf{I}_{p,q} \mathbf{X}_j \in [0, 1]$ for all $i, j \in [1, ..., n]$, where $\mathbf{I}_{p,q} = \text{diag}(1,...,1,-1,...,-1)$ with p ones and q minus ones on its diagonal. $p \geq 0$, $q \geq 1$, and $p + q = d$. $\mathbf{X} := [\mathbf{X}_1 | ... | \mathbf{X}_n]^T \in \mathbb{R}^{n \times d}$. Let $\mathcal{F}$ be a joint distribution of $\mathbf{X}_1, ..., \mathbf{X}_n$ on $\mathbb{R}^d$. We say that $(\mathbf{X}, \mathbf{A}) \sim GRDPG(\mathcal{F})$ with signature (p, q) if the following hold. First, let $(\mathbf{X}_1, ..., \mathbf{X}_n) \sim \mathcal{F}$. Second, for all $i < j$, $\mathbf{A}_{ij} \overset{ind}{\sim} Bernoulli(\mathbf{X}_i^T \mathbf{I}_{p,q} \mathbf{X}_j)$.

The adjcency matrices generated by GRDPG model are symmetric and hollow; otherwise stated, the graphs generated by GRDPG are undirected without self-loops. To be specific, $\mathbf{A}_{ii} = 0$ for $i \in [1, ..., n]$.

2.3 Spectral Adjacency Embedding

In the real world, obtaining the adjacency matrix $\mathbf{A}$ of the social network is straightforward. The problem people concern is how to recover latent positions from $\mathbf{A}$ upon RDPG model or GRDPG model. The common two methods to extract information of latent positions are Adjacency Spectral Embedding (ASE) and Laplacian Spectral Embedding (LSE). (Athreya et al., 2017) The one of our interest in this thesis is ASE.

Definition 2. (ASE). Given an undirected hollow graph $\mathbf{A}$ and a positive number $d \geq 1$, the adjacency spectral embedding (ASE) of $\mathbf{A}$ into $\mathbb{R}^d$ is $\hat{\mathbf{X}} = \mathbf{U}_\mathbf{A} \mathbf{S}_\mathbf{A}^{1/2}$, where $\mathbf{S}_\mathbf{A}$ is the diagonal matrix of square root of d largest eigenvalues of $\mathbf{A}^T\mathbf{A}$ and $\mathbf{U}_\mathbf{A} \in \mathbb{R}^d$ is the matrix whose columns are corresponding eigenvectors.

$\hat{\mathbf{X}}$, the results of ASE, is a good estimate of $\mathbf{X}$. The principle behind is as follows. $\mathbf{A}$ can be viewed as a perturbation of $\mathbf{P}$. Namely $\mathbf{A} = \mathbf{P} + \mathbf{E}$, where $\mathbf{E}$ is a noise matrix. When noise is relatively small, the leading eigenvalues and eigenvectors of $\mathbf{A}$ and $\mathbf{P}$ should be close. Let $\mathbf{A} = \mathbf{U}_\mathbf{A} \mathbf{\Sigma}_\mathbf{A} \mathbf{U}_\mathbf{A}^T$ be the eigendecomposition of $\mathbf{A}$. Then $\mathbf{A} = \mathbf{U}_\mathbf{A} \mathbf{\Sigma}_\mathbf{A} \mathbf{U}_\mathbf{A}^T = \mathbf{U}_\mathbf{A} |\mathbf{\Sigma}_\mathbf{A}|^{\frac{1}{2}} \mathbf{I}_{p,q} |\mathbf{\Sigma}_\mathbf{A}|^{\frac{1}{2}} \mathbf{U}_\mathbf{A}^T$, where $|\mathbf{\Sigma}_\mathbf{A}| \in \mathbb{R}^{d \times d}$ takes entrywise absolute values of $\mathbf{\Sigma}_\mathbf{A}$, in which the diagonal contains d largest eigenvalues of $\mathbf{A}$ in magnitude. the diagonal entries of $\mathbf{I}_{p,q}$

take values of 1 if their corresponding eigenvalues are positive and take values of -1 if their corresponding eigenvalues are negative. As a deduction, if we use GRDPG model in which $\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^T$, $\mathbf{U_A}|\mathbf{\Sigma_A}|^{\frac{1}{2}} = \mathbf{U_A}\mathbf{S_A}^{1/2}$ should be close to $\mathbf{X}$. Furthermore, if $\mathbf{P}$ is positive semidefinite, then $p = d$, $q = 0$, and thus $\mathbf{I}_{p,q}$ becomes an indentity matrix. Using RDPG model in which $\mathbf{P} = \mathbf{X}\mathbf{X}^T$, $\mathbf{U_A}\mathbf{S_A}^{1/2}$ should be still close to $\mathbf{X}$.

2.4 Methodology

For our problem, assume we can observe the adjacency matrix $\mathbf{A}$ where $\mathbf{A}_{ij} \in \{0, 1\}$ indicates whether there is connection between vertex i and j, and a covariates matrix $\mathbf{Z} \in \mathbb{R}^{n \times m}$ where $\mathbf{Z}_i \in \mathbb{R}^m$ records m covariates of observation i. To model such network data, we build GRDPG with covariates as follows.

$$\mathbf{A}_{ij}|\mathbf{X}_i, \mathbf{X}_j \overset{ind}{\sim} Bernoulli(\mathbf{X}_i^T \mathbf{I}_{p,q} \mathbf{X}_j + f(\mathbf{Z}_i, \mathbf{Z}_j; \boldsymbol{\beta}))$$

where f is a known function, $\boldsymbol{\beta}$ is vector of covariates parameters, $\mathbf{X}_i \in \mathbb{R}^d$ is the latent position of vertex i with $i \in \{0, ..., n\}$, $\mathbf{I}_{p,q}$ is a diagonal matrix with p ones and q minus ones ($p + q = d$) on its diagonal. Our goal is to estimate $\boldsymbol{\beta}$ and $\mathbf{X}$, where $\mathbf{X} = [\mathbf{X}_1|...|\mathbf{X}_n]^T \in \mathbb{R}^{n \times d}$. In fact, once an accurate estimate of $\boldsymbol{\beta}$ is found, we can remove the effects of covariates and then use the ASE discussed in 2.3 to estimate $\mathbf{X}$. (Lyzinski et al., 2017; Lyzinski et al., 2014; Athreya et al., 2016; Sussman et al., 2012; Tang et al., 2017) Moreover, we can cluster communities of $\hat{X}$ using Gaussian Mixture Modelling (GMM).(Mele et al., 2019) Thus, we focus on estimating $\boldsymbol{\beta}$ in this thesis.

2.4.1 GRDPG with one covariate

Assume $m = 1$ and $Z_i \in \{-1, 1\}$. Then $\mathbf{Z}$ is a covariate vector of length n. We build the following model.

$$\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^T + \beta\frac{\mathbf{1}\mathbf{1}^T + \mathbf{Z}\mathbf{Z}^T}{2}$$

$$\mathbf{A}_{ij}|\mathbf{X}_i, \mathbf{X}_j \overset{ind}{\sim} Bernoulli(\mathbf{P}_{ij})$$

where $\mathbf{1}$ is the column vector of length n whose entries are all 1's. Note that

$$\begin{aligned}\mathbf{P} &= \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^T + \beta\frac{\mathbf{1}\mathbf{1}^T + \mathbf{Z}\mathbf{Z}^T}{2} \\ &= \begin{bmatrix} \mathbf{X}_1\mathbf{I}_{p,q}\mathbf{X}_1^T + \beta\mathbf{1}\mathbf{1}^T & \mathbf{X}_1\mathbf{I}_{p,q}\mathbf{X}_{-1}^T \\ \mathbf{X}_{-1}\mathbf{I}_{p,q}\mathbf{X}_1^T & \mathbf{X}_{-1}\mathbf{I}_{p,q}\mathbf{X}_{-1}^T + \beta\mathbf{1}\mathbf{1}^T \end{bmatrix}\end{aligned}$$

where $\mathbf{X}_1$ denotes those vertices with covariates equal to 1, and $\mathbf{X}_{-1}$ denotes those vertices with covariates equal to -1.

Applying ASE on the adjacency matrix $\mathbf{A}$ into $\mathbb{R}^{d+2}$, we have $\mathbf{Y} := \mathbf{U}_\mathbf{A}\mathbf{S}_\mathbf{A}^{1/2} \in \mathbb{R}^{n\times(d+2)}$ such that $\mathbf{P} \simeq \hat{\mathbf{P}} = \mathbf{Y}\mathbf{I}_{p,q}\mathbf{Y}^T$, where $\mathbf{I}_{p,q}$ is a diagonal matrix whose diagonal entries take values of 1 if corresponding eigenvalues of $\mathbf{A}$ is positive and take values of -1 if corresponding eigenvalues of $\mathbf{A}$ is negative. $\hat{\mathbf{P}}$ should have the following structure:

$$\begin{aligned}\hat{\mathbf{P}} &= \hat{\mathbf{X}}\mathbf{I}_{p,q}\hat{\mathbf{X}}^T + \hat{\beta}\frac{\mathbf{1}\mathbf{1}^T + \mathbf{Z}\mathbf{Z}^T}{2} \\ &= \begin{bmatrix} \hat{\mathbf{X}}_1\mathbf{I}_{p,q}\hat{\mathbf{X}}_1^T + \hat{\beta}\mathbf{1}\mathbf{1}^T & \hat{\mathbf{X}}_1\mathbf{I}_{p,q}\hat{\mathbf{X}}_{-1}^T \\ \hat{\mathbf{X}}_{-1}\mathbf{I}_{p,q}\hat{\mathbf{X}}_1^T & \hat{\mathbf{X}}_{-1}\mathbf{I}_{p,q}\hat{\mathbf{X}}_{-1}^T + \hat{\beta}\mathbf{1}\mathbf{1}^T \end{bmatrix}\end{aligned}$$

We now introduce some notations:

$n_1 := |\{i : Z_i = 1\}|$

$n_{-1} := |\{i : Z_i = -1\}|$

$n = n_1 + n_{-1}$

Let $\mathbf{J}_1$ be the centering matrix $\mathbf{I} - n_1^{-1}\mathbf{1}\mathbf{1}^T$, $\mathbf{J}_{-1} = \mathbf{I} - n_{-1}^{-1}\mathbf{1}\mathbf{1}^T$

Center $\hat{\mathbf{P}}$:

$$\begin{aligned}
\tilde{\mathbf{H}} &:= \begin{bmatrix} \mathbf{J}_1(\hat{\mathbf{X}}_1\mathbf{I}_{p,q}\hat{\mathbf{X}}_1^T + \hat{\beta}\mathbf{1}\mathbf{1}^T)\mathbf{J}_1 & \mathbf{J}_1(\hat{\mathbf{X}}_1\mathbf{I}_{p,q}\hat{\mathbf{X}}_{-1}^T)\mathbf{J}_{-1} \\ \mathbf{J}_{-1}(\hat{\mathbf{X}}_{-1}\mathbf{I}_{p,q}\hat{\mathbf{X}}_1^T)\mathbf{J}_1 & \mathbf{J}_{-1}(\hat{\mathbf{X}}_{-1}\mathbf{I}_{p,q}\hat{\mathbf{X}}_{-1}^T + \hat{\beta}\mathbf{1}\mathbf{1}^T)\mathbf{J}_{-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{J}_1\hat{\mathbf{X}}_1\mathbf{I}_{p,q}\hat{\mathbf{X}}_1^T\mathbf{J}_1 & \mathbf{J}_1\hat{\mathbf{X}}_1\mathbf{I}_{p,q}\hat{\mathbf{X}}_{-1}^T\mathbf{J}_{-1} \\ \mathbf{J}_{-1}\hat{\mathbf{X}}_{-1}\mathbf{I}_{p,q}\hat{\mathbf{X}}_1^T\mathbf{J}_1 & \mathbf{J}_{-1}\hat{\mathbf{X}}_{-1}\mathbf{I}_{p,q}\hat{\mathbf{X}}_{-1}^T\mathbf{J}_{-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{J}_1\hat{\mathbf{X}}_1 \\ \mathbf{J}_{-1}\hat{\mathbf{X}}_{-1} \end{bmatrix} \mathbf{I}_{p,q} \begin{bmatrix} \mathbf{J}_1\hat{\mathbf{X}}_1 \\ \mathbf{J}_{-1}\hat{\mathbf{X}}_{-1} \end{bmatrix}^T
\end{aligned}$$

Accordingly, $\mathbf{H} := \begin{bmatrix} \mathbf{J}_1\mathbf{X}_1\mathbf{I}_{p,q}\mathbf{X}_1^T\mathbf{J}_1 & \mathbf{J}_1\mathbf{X}_1\mathbf{I}_{p,q}\mathbf{X}_{-1}^T\mathbf{J}_{-1} \\ \mathbf{J}_{-1}\mathbf{X}_{-1}\mathbf{I}_{p,q}\mathbf{X}_1^T\mathbf{J}_1 & \mathbf{J}_{-1}\mathbf{X}_{-1}\mathbf{I}_{p,q}\mathbf{X}_{-1}^T\mathbf{J}_{-1} \end{bmatrix}$

Compute SVD (apply ASE) of the top-left block of $\tilde{\mathbf{H}}$, that is $\mathbf{U}_1\mathbf{\Sigma}_1\mathbf{U}_1^T$. Define $\tilde{\mathbf{Y}}_2 := \mathbf{U}_1\mathbf{\Sigma}_1^{\frac{1}{2}} \in \mathbb{R}^{n_1 \times d}$, $\mathbf{Y}_2 := \begin{bmatrix} \tilde{\mathbf{Y}}_2 & \mathbf{0} \end{bmatrix}$. Then we have

$$\mathbf{Y}_2 \simeq \begin{bmatrix} \mathbf{X}_1 - \boldsymbol{\mu}_1 & \mathbf{0} \end{bmatrix} \tag{2.1}$$

On the other hand,

$$\hat{\mathbf{X}}_1\mathbf{I}_{p,q}\hat{\mathbf{X}}_1^T + \hat{\beta}\mathbf{1}\mathbf{1}^T = \begin{bmatrix} \hat{\mathbf{X}}_1 & \sqrt{\hat{\beta}}\mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{p,q} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{X}}_1 & \sqrt{\hat{\beta}}\mathbf{1} \end{bmatrix}^T$$

Applying ASE on the top left block of matrix $\mathbf{A}$, we have $\mathbf{Y}_1 \in \mathbb{R}^{n_1 \times (d+1)}$ such that $\mathbf{X}_1 \mathbf{I}_{p,q} \mathbf{X}_1^T + \beta \mathbf{1}\mathbf{1}^T \simeq \mathbf{Y}_1 \begin{bmatrix} \mathbf{I}_{p,q} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \mathbf{Y}_1^T$. Therefore,

$$\mathbf{Y}_1 \simeq \begin{bmatrix} \mathbf{X}_1 & \sqrt{\beta}\mathbf{1} \end{bmatrix} \tag{2.2}$$

(2.2) minus (2.1) implies

$$\mathbf{Y}_1 - \mathbf{Y}_2 \simeq \begin{bmatrix} \boldsymbol{\mu}_1 & \sqrt{\beta}\mathbf{1} \end{bmatrix} \tag{2.3}$$

(2.3) shows the last column of $\mathbf{Y}_1 - \mathbf{Y}_2$ can be used to estimate $\sqrt{\beta}$ and hence estimate β. From that, define $\mathbf{u}_1 \in \mathbb{R}^{n_1}$ to be the last column of $\mathbf{Y}_1 - \mathbf{Y}_2$, then the estimate of β is

$$\hat{\beta} := \frac{1}{n_1} ||\mathbf{u}_1||_2^2$$

On top of that, Chapter 4 will rigorously prove that the last column of $\mathbf{Y}_1 - \mathbf{Y}_2$ is asymptotically equal to $\sqrt{\beta}\mathbf{1}$ up to orthogonal transformation. Chapter 5 will present a specific procedure to find those orthogonal transformation in practice.

2.4.2 GRDPG with multiple independent covariates

Assume that there are two independent covariates, both of which take value from $\{-1, 1\}$. Namely $\mathbf{Z} \in \mathbb{R}^{n \times 2}$ and $\mathbf{Z}_i \in \{-1, 1\}^2$. The model of GRDPG with covariates in this case is

$$\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^T + \beta_1 \frac{\mathbf{1}\mathbf{1}^T + \mathbf{Z}^{(1)}\mathbf{Z}^{(1)T}}{2} + \beta_2 \frac{\mathbf{1}\mathbf{1}^T + \mathbf{Z}^{(2)}\mathbf{Z}^{(2)T}}{2}$$

$$\mathbf{A}_{ij} | \mathbf{X}_i, \mathbf{X}_j \overset{ind}{\sim} Bernoulli(\mathbf{P}_{ij})$$

where $\mathbf{Z}^{(i)}$ is the i^{th} column of $\mathbf{Z}$.

$$\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^T + \beta_1 \frac{\mathbf{1}\mathbf{1}^T + \mathbf{Z}^{(1)}\mathbf{Z}^{(1)T}}{2} + \beta_2 \frac{\mathbf{1}\mathbf{1}^T + \mathbf{Z}^{(2)}\mathbf{Z}^{(2)T}}{2}$$

$$= \begin{bmatrix} \mathbf{X}_{1,1}\mathbf{I}_{p,q}\mathbf{X}_{1,1}^T + \beta_1\mathbf{1}\mathbf{1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \mathbf{X}_{1,1}\mathbf{I}_{p,q}\mathbf{X}_{1,-1}^T + \beta_1\mathbf{1}\mathbf{1}^T & \boxed{\mathbf{X}_{1,1}\mathbf{I}_{p,q}\mathbf{X}_{-1,1}^T + \beta_2\mathbf{1}\mathbf{1}^T} & \boxed{\mathbf{X}_{1,1}\mathbf{I}_{p,q}\mathbf{X}_{-1,-1}^T} \\ \mathbf{X}_{1,-1}\mathbf{I}_{p,q}\mathbf{X}_{1,1}^T + \beta_1\mathbf{1}\mathbf{1}^T & \mathbf{X}_{1,-1}\mathbf{I}_{p,q}\mathbf{X}_{1,-1}^T + \beta_1\mathbf{1}\mathbf{1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \boxed{\mathbf{X}_{1,-1}\mathbf{I}_{p,q}\mathbf{X}_{-1,1}^T} & \boxed{\mathbf{X}_{1,-1}\mathbf{I}_{p,q}\mathbf{X}_{-1,-1}^T + \beta_2\mathbf{1}\mathbf{1}^T} \\ \mathbf{X}_{-1,1}\mathbf{I}_{p,q}\mathbf{X}_{1,1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \mathbf{X}_{-1,1}\mathbf{I}_{p,q}\mathbf{X}_{1,-1}^T & \mathbf{X}_{-1,1}\mathbf{I}_{p,q}\mathbf{X}_{-1,1}^T + \beta_1\mathbf{1}\mathbf{1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \mathbf{X}_{-1,1}\mathbf{I}_{p,q}\mathbf{X}_{-1,-1}^T + \beta_1\mathbf{1}\mathbf{1}^T \\ \mathbf{X}_{-1,-1}\mathbf{I}_{p,q}\mathbf{X}_{1,1}^T & \mathbf{X}_{-1,-1}\mathbf{I}_{p,q}\mathbf{X}_{1,-1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \mathbf{X}_{-1,-1}\mathbf{I}_{p,q}\mathbf{X}_{-1,1}^T + \beta_1\mathbf{1}\mathbf{1}^T & \mathbf{X}_{-1,-1}\mathbf{I}_{p,q}\mathbf{X}_{-1,-1}^T + \beta_1\mathbf{1}\mathbf{1}^T + \beta_2\mathbf{1}\mathbf{1}^T \end{bmatrix} \tag{2.4}$$

$$= \begin{bmatrix} \mathbf{X}_{1,1}\mathbf{I}_{p,q}\mathbf{X}_{1,1}^T + \beta_1\mathbf{1}\mathbf{1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \boxed{\mathbf{X}_{1,1}\mathbf{I}_{p,q}\mathbf{X}_{1,-1}^T + \beta_1\mathbf{1}\mathbf{1}^T} & \mathbf{X}_{1,1}\mathbf{I}_{p,q}\mathbf{X}_{-1,1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \boxed{\mathbf{X}_{1,1}\mathbf{I}_{p,q}\mathbf{X}_{-1,-1}^T} \\ \mathbf{X}_{1,-1}\mathbf{I}_{p,q}\mathbf{X}_{1,1}^T + \beta_1\mathbf{1}\mathbf{1}^T & \mathbf{X}_{1,-1}\mathbf{I}_{p,q}\mathbf{X}_{1,-1}^T + \beta_1\mathbf{1}\mathbf{1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \mathbf{X}_{1,-1}\mathbf{I}_{p,q}\mathbf{X}_{-1,1}^T & \mathbf{X}_{1,-1}\mathbf{I}_{p,q}\mathbf{X}_{-1,-1}^T + \beta_2\mathbf{1}\mathbf{1}^T \\ \mathbf{X}_{-1,1}\mathbf{I}_{p,q}\mathbf{X}_{1,1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \boxed{\mathbf{X}_{-1,1}\mathbf{I}_{p,q}\mathbf{X}_{1,-1}^T} & \mathbf{X}_{-1,1}\mathbf{I}_{p,q}\mathbf{X}_{-1,1}^T + \beta_1\mathbf{1}\mathbf{1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \boxed{\mathbf{X}_{-1,1}\mathbf{I}_{p,q}\mathbf{X}_{-1,-1}^T + \beta_1\mathbf{1}\mathbf{1}^T} \\ \mathbf{X}_{-1,-1}\mathbf{I}_{p,q}\mathbf{X}_{1,1}^T & \mathbf{X}_{-1,-1}\mathbf{I}_{p,q}\mathbf{X}_{1,-1}^T + \beta_2\mathbf{1}\mathbf{1}^T & \mathbf{X}_{-1,-1}\mathbf{I}_{p,q}\mathbf{X}_{-1,1}^T + \beta_1\mathbf{1}\mathbf{1}^T & \mathbf{X}_{-1,-1}\mathbf{I}_{p,q}\mathbf{X}_{-1,-1}^T + \beta_1\mathbf{1}\mathbf{1}^T + \beta_2\mathbf{1}\mathbf{1}^T \end{bmatrix} \tag{2.5}$$

where $\mathbf{X}_{i,j} \in \mathbb{R}^{n_{i,j} \times d}$ denotes the latent position matrix of those vertices with first covariate equal to $i \in \{-1, 1\}$ and second covariate equal to $j \in \{-1, 1\}$. $n_{i,j} := |\{l : \mathbf{Z}_l^{(1)} = i, \mathbf{Z}_l^{(2)} = j\}|$.

The underlying structure of $\mathbf{P}$ is shown above. Four circled blocks in (2.4) can form a new matrix that can be used to estimate β_2 through the methodology described in Section 2.4. Similarly, four circled blocks in (2.5) can be used to estimate β_1. Note that some permutation in terms of covariates need to be operated to make $\mathbf{A}$ correspond to matrix (2.4).

For number of independent covariates greater than 2, $\mathbf{A}$ can be permuted to be aligned with a similar structure of (2.4). Then $\boldsymbol{\beta}$ can be estimated by building new adjacency matrices for each parameter in $\boldsymbol{\beta}$ based on sub-blocks of $\mathbf{A}$.

References

Holland, P.W., K.B. Laskey, and S. Leinhardt (1983). "Stochastic Blockmodels: First Steps". In: *Social Networks* 5.2, pp. 109–137.

Diaconis, P. and S. Janson (2007). "Graph limits and exchangeable random graphs". In: *arXiv preprint arXiv:0712.2749*.

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). "Latent space approaches to social network analysis". In: *Journal of the American Statistical Association* 97, pp. 1090–1098.

Smith, A. L., D. Asta, and C. A. Calder (2017). "The geometry of continuous latent space models for network data". In: *Arxiv preprint arxiv:1712.08641*.

Young, S. and E. Scheinerman (2007). "Random dot product graph models for social networks". In: *Proceedings of the 5th international conference on algorithms and models for the web-graph*, pp. 138–149.

Athreya, A., D. E. Fishkind, K. Levin, V. Lyzinski, Y. Park, Y. Qin, D. L. Sussman, M. Tang, J. T. Vogelstein, and C. E. Priebe (2017). "Statistical inference on random dot product graphs: a survey". In: *arXiv preprint arXiv:1709.05454*.

Rubin-Delanchy, P., C. E. Priebe, and M. Tang (2017). "A statistical interpretation of spectral embedding: the generalised random dot product graph". In: *arXiv:1709.05506*.

Lyzinski, V., M. Tang, A. Athreya, Y. Park, and C. E. Priebe (2017). "Community detection and classification in hierarchical stochastic blockmodels". In: *IEEE Transactions in Network Science and Engineering* 4, pp. 13–16.

Lyzinski, V., D. L. Sussman, M. Tang, A. Athreya, and C. E. Priebe (2014). "Perfect clustering for stochastic blockmodel graphs via adjacency spectral embedding". In: *Electronic Journal of Statistics* 8, pp. 2905–2922.

Athreya, A., V. Lyzinski, D. J. Marchette, C. E. Priebe, D. L. Sussman, and M. Tang (2016). "A limit theorem for scaled eigenvectors of random dot product graphs". In: *Sankhya A* 78, pp. 1–18.

Sussman, D. L., M. Tang, D. E. Fishkind, and C. E. Priebe (2012). "A consistent adjacency spectral embedding for stochastic blockmodel graphs". In: *Journal of the American Statistical Association* 107, pp. 1119–1128.
Tang, M., A. Athreya, D. L. Sussman, V. Lyzinski, and C. E. Priebe (2017). "A nonparametric two-sample hypothesis testing problem for random dot product graphs". In: *Bernoulli* 23, pp. 1599–1630.
Mele, A., L. Hao, J. Cape, and C.E. Priebe (2019). "Spectral inference for large stochastic blockmodels with nodal covariates". In: *arXiv:1908.06438.*

Chapter 3

Main theoretical Results

Following the same notations and definitions in Chapter 2, this chapter will prove the consistency of the estimation of $\boldsymbol{\beta}$ proposed in Section 2.4. First consider GRDPG with one binary covariates. We now prove that $\hat{\beta} = n_1^{-1}||\mathbf{u}_1||_2^2$ is asymptotically equal to β up to orthogonal transformation.

Proof. Initially prove

$$\mathbf{Y}_2 - \begin{bmatrix} \mathbf{J}_1\mathbf{X}_1 & \mathbf{0} \end{bmatrix} \to \mathbf{0}$$

Lemma 1 (Tang et al., 2013)**.** Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ positive semidefinite matrices with rank($\mathbf{A}$)=rank($\mathbf{B}$)=d. Let $\mathbf{X}, \mathbf{Y} \in \mathcal{M}_{n,d}(\mathbb{R})$ be of full column rank such that $\mathbf{X}\mathbf{X}^T = \mathbf{A}$ and $\mathbf{Y}\mathbf{Y}^T = \mathbf{B}$. Let δ be the smallest nonzero eigenvalue of $\mathbf{B}$. Then there exists an orthogonal matrix $\mathbf{W} \in \mathcal{M}_d(\mathbb{R})$ such that

$$||\mathbf{X}\mathbf{W} - \mathbf{Y}||_F \leq \frac{||\mathbf{A} - \mathbf{B}||(\sqrt{d||\mathbf{A}||} + \sqrt{d||\mathbf{B}||})}{\delta}$$

The proof of **Lemma 1** can be found in Appendix.

Recall that we compute SVD of the top-left block of $\tilde{\mathbf{H}}$, that is $\mathbf{U}_1\mathbf{\Sigma}_1\mathbf{U}_1^T$. Define $\tilde{\mathbf{Y}}_2 := \mathbf{U}_1\mathbf{\Sigma}_1^{\frac{1}{2}} \in \mathbb{R}^{n_1\times d}$. In such way $\tilde{\mathbf{Y}}_2$ is ASE of $\tilde{\mathbf{H}}$. Also define $\mathbf{Y}_2 := \begin{bmatrix}\tilde{\mathbf{Y}}_2 & \mathbf{0}\end{bmatrix}$. Let δ_3 be the smallest nonzero eigenvalue of $\mathbf{J}_1\mathbf{X}_1\mathbf{X}_1^T\mathbf{J}_1$. Then there exists an orthogonal matrix $\tilde{\mathbf{T}}_3 \in \mathbb{R}^{d\times d}$ and $\mathbf{T}_3 = \begin{bmatrix}\tilde{\mathbf{T}}_3 & \mathbf{0} \\ \mathbf{0} & 1\end{bmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}$ such that

$$
\begin{aligned}
&||\mathbf{Y}_2\mathbf{T}_3 - \begin{bmatrix}\mathbf{J}_1\mathbf{X}_1 & \mathbf{0}\end{bmatrix}||_F && (3.1)\\
=&||\tilde{\mathbf{Y}}_2\tilde{\mathbf{T}}_3 - \mathbf{J}_1\mathbf{X}_1||_F && (3.2)\\
=&||\mathbf{U}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\tilde{\mathbf{T}}_3 - \mathbf{J}_1\mathbf{X}_1||_F && (3.3)\\
\leq&\frac{\sqrt{d}||\mathbf{U}_1\mathbf{\Sigma}_1\mathbf{U}_1^T - \mathbf{J}_1\mathbf{X}_1\mathbf{X}_1^T\mathbf{J}_1||(||\mathbf{U}_1\mathbf{\Sigma}_1^{\frac{1}{2}}|| + ||\mathbf{J}_1\mathbf{X}_1||)}{\delta_3} && (3.4)\\
\leq&\frac{\sqrt{d}||\mathbf{U}_1\mathbf{\Sigma}_1\mathbf{U}_1^T - \mathbf{J}_1\mathbf{X}_1\mathbf{X}_1^T\mathbf{J}_1||_F(||\mathbf{U}_1\mathbf{\Sigma}_1^{\frac{1}{2}}|| + ||\mathbf{J}_1\mathbf{X}_1||)}{\delta_3} && (3.5)\\
=&\frac{\sqrt{d}||\mathbf{U}_1\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{I}_{p,q}\mathbf{\Sigma}_1^{\frac{1}{2}}\mathbf{U}_1^T - \mathbf{J}_1\mathbf{X}_1\mathbf{I}_{p,q}\mathbf{X}_1^T\mathbf{J}_1||_F(||\mathbf{U}_1\mathbf{\Sigma}_1^{\frac{1}{2}}|| + ||\mathbf{J}_1\mathbf{X}_1||)}{\delta_3} && (3.6)\\
\leq&\frac{\sqrt{d}||\tilde{\mathbf{H}} - \mathbf{H}||_F(||\mathbf{U}_1\mathbf{\Sigma}_1^{\frac{1}{2}}|| + ||\mathbf{J}_1\mathbf{X}_1||)}{\delta_3} && (3.7)\\
\leq&\frac{\sqrt{d}||\mathbf{Y}\mathbf{I}_{p,q}\mathbf{Y}^T - \mathbf{P}||_F(||\mathbf{U}_1\mathbf{\Sigma}_1^{\frac{1}{2}}|| + ||\mathbf{J}_1\mathbf{X}_1||)}{\delta_3} && (3.8)\\
=&\frac{\sqrt{d}\cdot O(\sqrt{nlog^{d+2}n})\cdot O(\sqrt{n_1})}{O(n_1)} && (3.9)\\
=&O(\sqrt{\frac{nd}{n_1}\cdot log^{d+2}n}) && (3.10)
\end{aligned}
$$

(3.4) follows **lemma 1**. (3.5) follows the fact that the spectral norm is less than or equal to the Frobenius norm. (3.6) uses the fact that changing signs of eigenvalues simultaneously does not influence on the absolute value of the difference. (3.7) holds due to the fact that

$$||\mathbf{J}\mathbf{X}\mathbf{X}^T\mathbf{J}||_F \leq ||\mathbf{J}|| \cdot ||\mathbf{X}\mathbf{X}^T||_F||\mathbf{J}||$$

$$= ||\mathbf{X}\mathbf{X}^T||_F$$

where $\mathbf{J} := \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}^T$ is a centering matrix.

Remark. The spectral norm of centering matrix $\mathbf{J}$ is 1; namely, $||\mathbf{J}|| = 1$.

In (3.9), $||\mathbf{Y}\mathbf{I}_{p,q}\mathbf{Y}^T - \mathbf{P}||_F \leq O(\sqrt{nlog^{d+2}n})$ is given by Xu. (Xu, 2017)

Secondly prove

$$\mathbf{Y}_1 - \begin{bmatrix}\mathbf{X}_1 & \sqrt{\beta}\mathbf{1}\end{bmatrix} \to \mathbf{0}$$

Note that

$$\hat{\mathbf{X}}_1\mathbf{I}_{p,q}\hat{\mathbf{X}}_1^T + \hat{\beta}\mathbf{1}\mathbf{1}^T = \begin{bmatrix}\hat{\mathbf{X}}_1 & \sqrt{\hat{\beta}}\mathbf{1}\end{bmatrix} \begin{bmatrix}\mathbf{I}_{p,q} & \mathbf{0}\\ \mathbf{0} & 1\end{bmatrix} \begin{bmatrix}\hat{\mathbf{X}}_1 & \sqrt{\hat{\beta}}\mathbf{1}\end{bmatrix}^T$$

Applying ASE on the top left block of matrix $\mathbf{A}$, we have $\mathbf{Y}_1 \in \mathbb{R}^{n_1 \times (d+1)}$ such that $\mathbf{X}_1\mathbf{I}_{p,q}\mathbf{X}_1^T + \beta\mathbf{1}\mathbf{1}^T \simeq \mathbf{Y}_1 \begin{bmatrix}\mathbf{I}_{p,q} & \mathbf{0}\\ \mathbf{0} & 1\end{bmatrix} \mathbf{Y}_1^T$. Let δ_1 be the smallest eigenvalue of $(\mathbf{X}_1\mathbf{X}_1^T + \beta\mathbf{1}\mathbf{1}^T)$. Then there exists an orthogonal matrix $\mathbf{T}_1 \in \mathbb{R}^{(d+1)\times(d+1)}$ such that

$$||\mathbf{Y}_1\mathbf{T}_1 - \begin{bmatrix}\mathbf{X}_1 & \sqrt{\beta}\mathbf{1}\end{bmatrix}||_F \tag{3.11}$$

$$\leq \frac{\sqrt{d+1}||\mathbf{Y}_1\mathbf{Y}_1^T - (\mathbf{X}_1\mathbf{X}_1^T + \beta\mathbf{1}\mathbf{1}^T)||(||\mathbf{Y}_1|| + ||\begin{bmatrix}\mathbf{X}_1 & \sqrt{\beta}\mathbf{1}\end{bmatrix}||)}{\delta_1} \tag{3.12}$$

$$\leq \frac{\sqrt{d+1}||\mathbf{Y}_1\mathbf{Y}_1^T - (\mathbf{X}_1\mathbf{X}_1^T + \beta\mathbf{1}\mathbf{1}^T)||_F(||\mathbf{Y}_1|| + ||\begin{bmatrix}\mathbf{X}_1 & \sqrt{\beta}\mathbf{1}\end{bmatrix}||)}{\delta_1} \tag{3.13}$$

$$= \frac{\sqrt{d+1}||\mathbf{Y}_1\begin{bmatrix}\mathbf{I}_{p,q} & \mathbf{0}\\ \mathbf{0} & 1\end{bmatrix}\mathbf{Y}_1^T - (\mathbf{X}_1\mathbf{I}_{p,q}\mathbf{X}_1^T + \beta\mathbf{1}\mathbf{1}^T)||_F(||\mathbf{Y}_1|| + ||\begin{bmatrix}\mathbf{X}_1 & \sqrt{\beta}\mathbf{1}\end{bmatrix}||)}{\delta_1} \tag{3.14}$$

$$= \frac{\sqrt{d+1}\cdot O(\sqrt{n_1 log^{d+1} n_1})\cdot O(\sqrt{n_1})}{O(n_1)} \tag{3.15}$$

$$= O(\sqrt{(d+1)log^{d+1} n_1}) \tag{3.16}$$

Let $\mathbf{u}_2$ be the last column of $\mathbf{Y}_1\mathbf{T}_1 - \mathbf{Y}_2\mathbf{T}_2$. Next prove

$$\mathbf{u}_2 - \sqrt{\beta}\mathbf{1} \to \mathbf{0}$$

$$||\mathbf{u}_2 - \sqrt{\beta}\mathbf{1}||_F \tag{3.17}$$

$$\leq ||(\mathbf{Y}_1\mathbf{T}_1 - \mathbf{Y}_2\mathbf{T}_2) - \begin{bmatrix}\boldsymbol{\mu}_1 & \sqrt{\beta}\mathbf{1}\end{bmatrix}||_F \tag{3.18}$$

$$= ||(\mathbf{Y}_1\mathbf{T}_1 - \begin{bmatrix}\mathbf{X}_1 & \sqrt{\beta}\mathbf{1}\end{bmatrix}) + (\begin{bmatrix}\mathbf{X}_1 - \boldsymbol{\mu}_1 & \mathbf{0}\end{bmatrix} - \mathbf{Y}_2\mathbf{T}_2)||_F \tag{3.19}$$

$$\leq ||\mathbf{Y}_1\mathbf{T}_1 - \begin{bmatrix}\mathbf{X}_1 & \sqrt{\beta}\mathbf{1}\end{bmatrix}||_F + ||\mathbf{Y}_2\mathbf{T}_2 - \begin{bmatrix}\mathbf{X}_1 - \boldsymbol{\mu}_1 & \mathbf{0}\end{bmatrix}||_F \tag{3.20}$$

$$\leq O(\sqrt{(d+1)log^{d+1}n_1}) + O(\sqrt{\frac{nd}{n_1} \cdot log^{d+2}n}) \tag{3.21}$$

Define $\hat{\beta} := n_1^{-1}||\mathbf{u}_2||_2^2$, finally prove

$$\hat{\beta} - \beta \to 0$$

$$|\hat{\beta} - \beta| \tag{3.22}$$

$$= n_1^{-1}\left|||\mathbf{u}_2||^2 - ||\sqrt{\beta}\mathbf{1}||^2\right| \tag{3.23}$$

$$\leq n_1^{-1}\left|||\mathbf{u}_2|| - ||\sqrt{\beta}\mathbf{1}||\right| \cdot \left|||\mathbf{u}_2|| + ||\sqrt{\beta}\mathbf{1}||\right| \tag{3.24}$$

$$= n_1^{-1}\left|||\mathbf{u}_2|| - ||\sqrt{\beta}\mathbf{1}||\right| \cdot \left|||\mathbf{u}_2|| - ||\sqrt{\beta}\mathbf{1}|| + 2||\sqrt{\beta}\mathbf{1}||\right| \tag{3.25}$$

$$\leq n_1^{-1}\left|||\mathbf{u}_2|| - ||\sqrt{\beta}\mathbf{1}||\right| \cdot (\left|||\mathbf{u}_2|| - ||\sqrt{\beta}\mathbf{1}||\right| + 2||\sqrt{\beta}\mathbf{1}||) \tag{3.26}$$

$$\leq n_1^{-1}||\mathbf{u}_2 - \sqrt{\beta}\mathbf{1}|| \cdot (||\mathbf{u}_2 - \sqrt{\beta}\mathbf{1}|| + 2||\sqrt{\beta}\mathbf{1}||) \tag{3.27}$$

□

References

Tang, M., D. L. Sussman, and C. Priebe (2013). "Universally Consistent Vertex Classification For Latent Positions Graphs". In: *arXiv:1212.1182v3*.

Xu, Jiaming (2017). "Rates of Convergence of Spectral Methods for Graphon Estimation". In: *arXiv:1709.03183v1*.

Chapter 4

Simulations

4.1 Simulation Procedures

Chapter 3 has proved that $\mathbf{Y}_1 - \mathbf{Y}_2$ is asymptotically equal to $\sqrt{\beta}\mathbf{1}$ up to orthogonal transformation. Practically, we use Procrustes Procedures to exactly find those orthogonal transformation.

Consider GRDPG with one binary covariate, which is

$$\mathbf{P} = \mathbf{X}\mathbf{I}_{p,q}\mathbf{X}^T + \beta\frac{\mathbf{1}\mathbf{1}^T + \mathbf{Z}\mathbf{Z}^T}{2}$$

$$\mathbf{A}_{ij}|\mathbf{X}_i, \mathbf{X}_j \overset{ind}{\sim} Bernoulli(\mathbf{P}_{ij})$$

In practice, the algorithm to estimate β from $\mathbf{A}$ and $\mathbf{Z}$ is as follows:

(1) Compute ASE of $\mathbf{A}$; that is $\mathbf{Y}$

(2) Compute $\hat{\mathbf{P}} = \mathbf{Y}\mathbf{I}_{p,q}\mathbf{Y}^T$

(3) Permute $\hat{\mathbf{P}}$ in terms of covariates and center the top-left block of $\hat{\mathbf{P}}$, which records the connections between vertices with the same covariate value.

Denote the centered top-left block as $\mathbf{D}$

(4) Compute ASE of $\mathbf{D}$; that is $\tilde{\mathbf{Y}}_2$. $\mathbf{Y}_2 := \begin{bmatrix}\tilde{\mathbf{Y}}_2 & \mathbf{0}\end{bmatrix}$

(5) Compute ASE of the top-left block of $\mathbf{A}$; that is $\mathbf{Y}_1$

(6) Compute $\mathbf{C} = \mathbf{Y}_1^T \mathbf{J}_1 \mathbf{Y}_2$

(7) Compute the SVD of $\mathbf{C}$; that is, $\mathbf{C} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$

(8) The optimal rotation matrix is $\mathbf{W}_3 = \mathbf{V}\mathbf{U}^T$

(9) The optimal translation vector is $\mathbf{t}\mathbf{W}_2 = n_1^{-1}\mathbf{1}\mathbf{1}^T(\mathbf{Y}_1 - \mathbf{A}\mathbf{W}_3)$

(10) $\hat{\beta} = n_1^{-1}||\mathbf{u}||_2^2$, where $\mathbf{u}$ is the last column of $\mathbf{t}\mathbf{W}_2\mathbf{W}_3^T$

In step (4), $\tilde{\mathbf{Y}}_2$ is orthogonal transformation of $\mathbf{J}_1\tilde{\mathbf{X}}_1$ following the same notations in previous chapters. Namely, $\tilde{\mathbf{Y}}_2 = \mathbf{J}_1\tilde{\mathbf{X}}_1\mathbf{W}_1$ where $\mathbf{W}_1 \in \mathbb{R}^d$ is an orthogonal matrix. Let $\mathbf{W}_4 := \begin{bmatrix}\mathbf{W}_1 & \mathbf{0} \\ \mathbf{0} & 1\end{bmatrix} \in \mathbb{R}^{(d+1)\times(d+1)}$, then

$$\begin{aligned}
\mathbf{Y}_2 &= \begin{bmatrix}\mathbf{J}_1\hat{\mathbf{X}}_1\mathbf{W}_1 & \mathbf{0}\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{J}_1\hat{\mathbf{X}}_1 & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{W}_1 & \mathbf{0} \\ \mathbf{0} & 1\end{bmatrix} \\
&= \begin{bmatrix}\mathbf{J}_1\hat{\mathbf{X}}_1 & \mathbf{0}\end{bmatrix} \mathbf{W}_4
\end{aligned}$$

In step (5), $\mathbf{Y}_1$ is orthogonal transformation of $\begin{bmatrix}\hat{\mathbf{X}}_1 & \sqrt{\hat{\beta}}\mathbf{1}\end{bmatrix}$. Namely, $\mathbf{Y}_1 = \begin{bmatrix}\hat{\mathbf{X}}_1 & \sqrt{\hat{\beta}}\mathbf{1}\end{bmatrix}\mathbf{W}_2$, where $\mathbf{W}_2 \in \mathbb{R}^{(d+1)\times(d+1)}$ is an orthogonal matrix.

$$\begin{aligned}
\begin{bmatrix}\hat{\mathbf{X}}_1 & \sqrt{\hat{\beta}}\mathbf{1}\end{bmatrix}\mathbf{W}_2 &= (\begin{bmatrix}\hat{\mathbf{X}}_1 - \overline{\mathbf{X}}_1 & \mathbf{0}\end{bmatrix} + \begin{bmatrix}\overline{\mathbf{X}}_1 & \sqrt{\hat{\beta}}\mathbf{1}\end{bmatrix})\mathbf{W}_4\mathbf{W}_4^T\mathbf{W}_2 \\
&= \begin{bmatrix}\mathbf{J}_1\hat{\mathbf{X}}_1 & \mathbf{0}\end{bmatrix}\mathbf{W}_4\mathbf{W}_3 + \mathbf{t}\mathbf{W}_2
\end{aligned}$$

where $\mathbf{t} := \begin{bmatrix} \overline{\mathbf{X}}_1 & \sqrt{\hat{\beta}}\mathbf{1} \end{bmatrix}$, $\mathbf{W}_3 := \mathbf{W}_4^T\mathbf{W}_2$

Equivalently,

$$\mathbf{Y}_1 = \mathbf{Y}_2\mathbf{W}_3 + \mathbf{t}\mathbf{W}_2$$

Step (6) to step (9) follows Procrustes Procedures to find optimal rotation matrix $\mathbf{W}_3$ and optimal translation vector $\mathbf{t}\mathbf{W}_2$. (Borg et al., 2005) The detailed statement of Procrustean similarity transformations can be found in Appendix.

In step (10),

$$\mathbf{t}\mathbf{W}_2\mathbf{W}_3^T = \mathbf{t}\mathbf{W}_2\mathbf{W}_2^T\mathbf{W}_4$$

$$= \mathbf{t}\mathbf{W}_4$$

Due to the special structure of $\mathbf{W}_4$, the last column of $\mathbf{t}\mathbf{W}_4$ is the same as the last column of $\mathbf{t}$, which is $\sqrt{\hat{\beta}}\mathbf{1}$. Thus, the last column of $\mathbf{t}\mathbf{W}_2\mathbf{W}_3^T$ is equal to $\sqrt{\hat{\beta}}\mathbf{1}$.

4.2 Simulation Results

Consider that there are two communities in $\mathbf{X} \in \mathbb{R}^{n \times d}$. Let $\mathbf{p} \in \mathbb{R}^d$ denote the latent position of the first community and $\mathbf{q} \in \mathbb{R}^d$ denote the latent position of the second community. We investigate the difference between β and $\hat{\beta}$ by altering the value of β (Table 4.1 and Figure 4.1), the number of observations (Table 4.2 and Figure 4.2), and the value of d (Table 4.3 and Figure 4.3).

β	$\hat{\beta}$	$\|\beta - \hat{\beta}\|$	CPU Time (s)
0.1	0.08461645	0.015383553	17.61
0.15	0.15430248	0.004302476	14.10
0.2	0.20767166	0.007671655	11.32
0.25	0.25946646	0.009466464	11.59
0.3	0.31132318	0.011323182	12.10
0.35	0.36088746	0.010887459	12.11
0.4	0.41277985	0.012779847	13.51
0.45	0.46215600	0.012156002	12.39
0.5	0.51014223	0.010142233	11.01
0.55	0.55963482	0.009634821	11.15

Table 4.1: Test of β. n=4000, **p**=[0.2], **q**=[0.6], balanced block size and covariate size

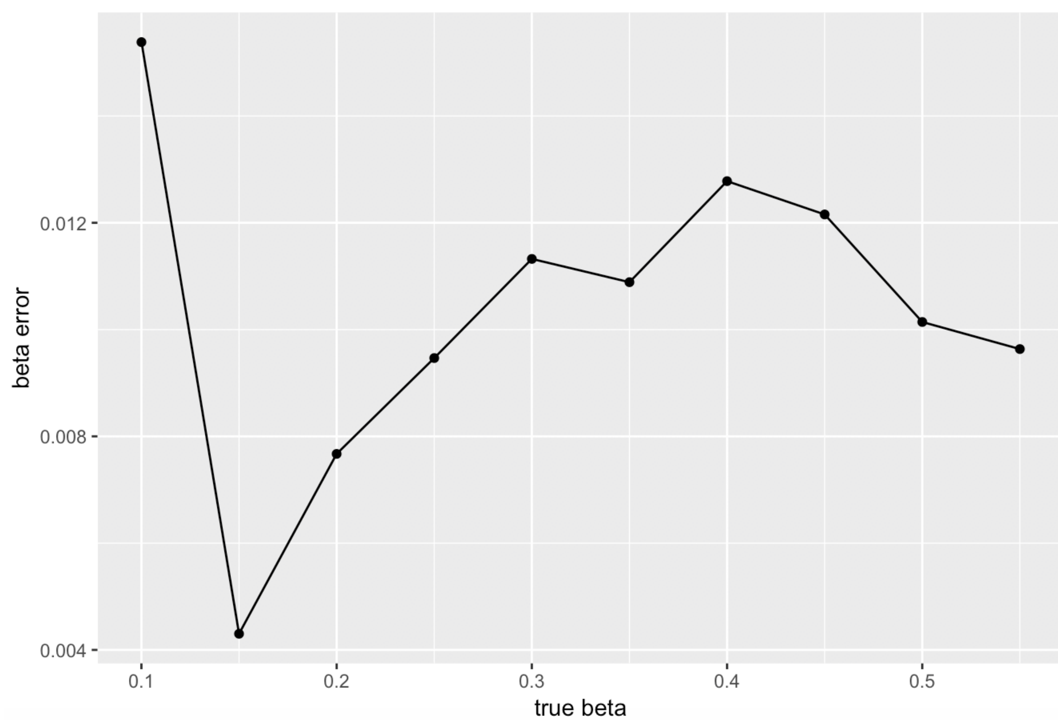

Figure 4.1: $|\beta - \hat{\beta}|$ **versus** β**.** n=4000, **p**=[0.2], **q**=[0.6], balanced block size and covariate size

n	β	$\hat{\beta}$	$\|\beta - \hat{\beta}\|$	CPU Time (s)
2000	0.35	0.3604602	0.010460235	1.504
4000	0.35	0.3608875	0.010887459	11.233
6000	0.35	0.3586939	0.008693881	39.861
8000	0.35	0.3559384	0.005938351	92.853
10000	0.35	0.3546021	0.004602101	180.533

Table 4.2: Test of n when $d = 1$. **p**=[0.2], **q**=[0.6], $\beta = 0.35$, balanced block size and covariate size

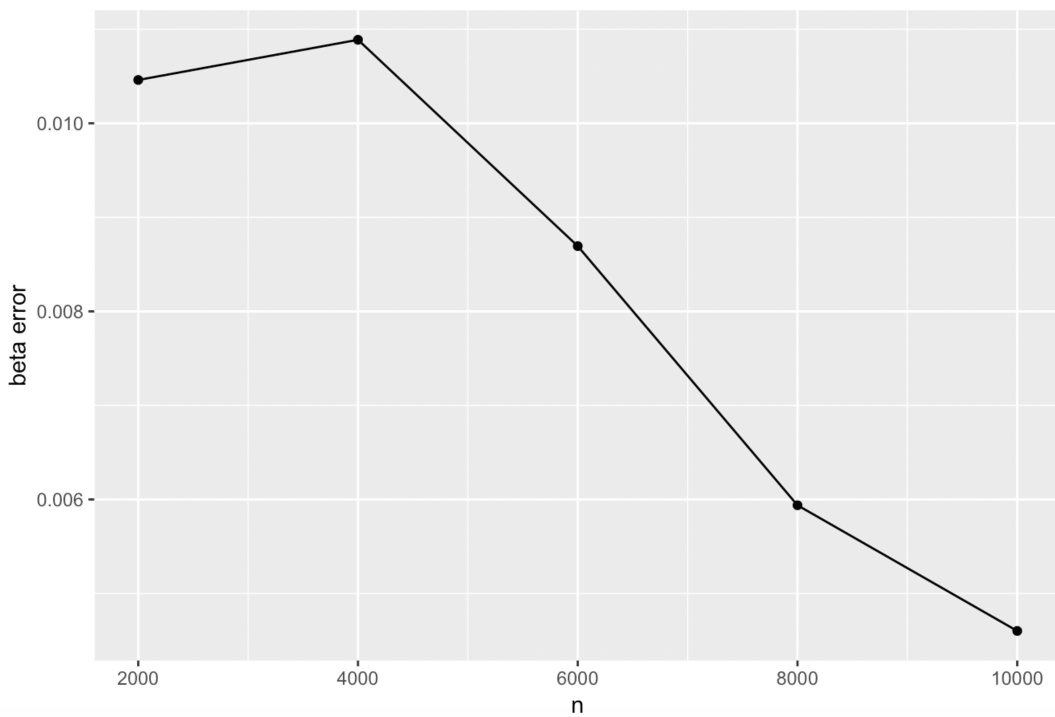

Figure 4.2: $|\beta - \hat{\beta}|$ versus n when $d = 1$. **p**=[0.2], **q**=[0.6], $\beta = 0.35$, balanced block size and covariate size

n	β	$\hat{\beta}$	$\lvert\beta - \hat{\beta}\rvert$	CPU Time (s)
2000	0.35	0.3852092	0.03520922	1.829
4000	0.35	0.3802840	0.03028395	15.172
6000	0.35	0.3775286	0.02752862	43.806
8000	0.35	0.3756289	0.02562886	105.502
10000	0.35	0.3741905	0.02419050	207.424

Table 4.3: Test of n when $d = 2$. $\mathbf{p} = [0.23, 0.14]^T$, $\mathbf{q} = [0.69, 0.13]^T$, $\beta = 0.35$, balanced block size and covariate size

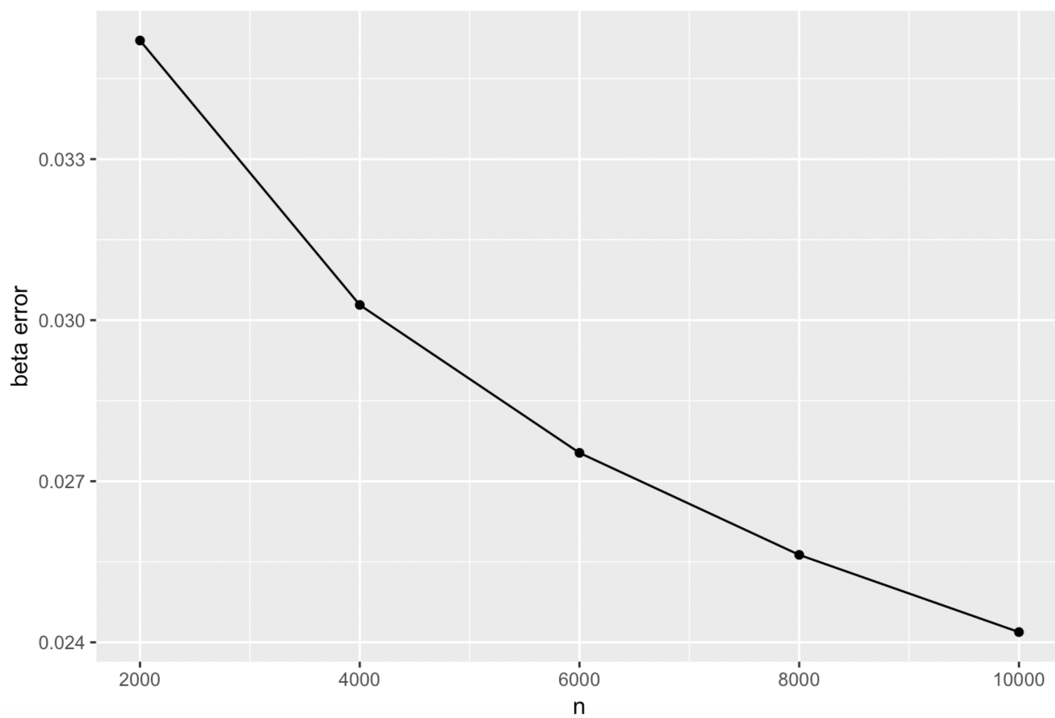

Figure 4.3: $\lvert\beta - \hat{\beta}\rvert$ **versus** n **when** $d = 2$**.** $\mathbf{p} = [0.23, 0.14]^T$, $\mathbf{q} = [0.69, 0.13]^T$, balanced block size and covariate size

As shown in Figure 4.2 and Figure 4.3, $\lvert\beta - \hat{\beta}\rvert$ decreases as n increases, which agrees the fact that $\mathbf{Y}_1 - \mathbf{Y}_2$ is asymptotically equal to $\sqrt{\beta}\mathbf{1}$ up to orthogonal transformation.

References

Borg, Ingwer and Patrick J. F. Groenen (2005). *Modern Multidimensional Scaling: Theory and Applications*. Springer New York, pp. 429–436.

Chapter 5

Conclusion and Future Work

We have successfully introduced *Generalized Random Dot Product graph with Covariates* model in matrix form to handle the network data with observed binary covariates. As the main contributions of this thesis, we have developed a spectral, Procrustes estimator for parameters in the model; furthermore, through both empirical tests and theoretical proof, we conclude that the estimator results are asymptotically equal to the true parameters up to orthogonal transformation.

In reality, both discrete covariates and continuous covariates can be observed. Take social network as an example, gender is a discrete covariate and income is a continuous covariate, both of which contribute to the probability of forming the link in a graph. Categorical covariates can be converted to binary covariates and then fit the model presented in this thesis. However, continuous covariates are out of discussion of this work. In future work, we expect to model graphs with continuous covariates and create an estimator for such models.

Chapter 6

Appendix

A.1

Lemma 1 (Tang et al., 2013)**.** Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ positive semidefinite matrices with rank($\mathbf{A}$)=rank($\mathbf{B}$)=d. Let $\mathbf{X}, \mathbf{Y} \in \mathcal{M}_{n,d}(\mathbb{R})$ be of full column rank such that $\mathbf{X}\mathbf{X}^T = \mathbf{A}$ and $\mathbf{Y}\mathbf{Y}^T = \mathbf{B}$. Let δ be the smallest nonzero eigenvalue of $\mathbf{B}$. Then there exists an orthogonal matrix $\mathbf{W} \in \mathcal{M}_d(\mathbb{R})$ such that

$$||\mathbf{X}\mathbf{W} - \mathbf{Y}||_F \leq \frac{||\mathbf{A} - \mathbf{B}||(\sqrt{d}||\mathbf{A}|| + \sqrt{d}||\mathbf{B}||)}{\delta}$$

Proof. Let $\mathbf{R} = \mathbf{A} - \mathbf{B}$. As $\mathbf{Y}$ is of full column rank, $\mathbf{Y}^T\mathbf{Y}$ is invertible, and its smallest eigenvalue is δ. We then have

$$\mathbf{Y} = \mathbf{X}\mathbf{X}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1} - \mathbf{R}\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}$$

Let $\mathbf{T} = \mathbf{X}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}$. We then have

$$\mathbf{T}^T\mathbf{T} - \mathbf{I} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{X}\mathbf{X}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1} - \mathbf{I} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{R}\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}$$

Therefore,

$$-(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T||\mathbf{R}||\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1} \preceq \mathbf{T}^T\mathbf{T} - \mathbf{I} \preceq (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T||\mathbf{R}||\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}$$

where $\preceq$ refers to the positive semi-definite ordering for matrices. We thus have

$$||\mathbf{T}^T\mathbf{T} - \mathbf{I}||_F \leq ||\mathbf{R}|| \cdot ||(\mathbf{Y}^T\mathbf{Y})^{-1}|_F \leq \sqrt{d}||\mathbf{R}|| \cdot ||(\mathbf{Y}^T\mathbf{Y})^{-1}| \leq \frac{||\mathbf{R}||\sqrt{d}}{\delta}$$

Now let $\mathbf{W}$ be the orthogonal matrix in the polar decomposition $\mathbf{T} = \mathbf{W}(\mathbf{T}^T\mathbf{T})^{1/2}$. We then have

$$\begin{aligned}
||\mathbf{XW} - \mathbf{Y}||_F &\leq ||\mathbf{XW} - \mathbf{XT}||_F + ||\mathbf{XT} - \mathbf{Y}||_F \\
&\leq ||\mathbf{X}|| \cdot ||(\mathbf{T}^T\mathbf{T})^{1/2} - \mathbf{I}||_F + ||\mathbf{R}|| \cdot ||\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}||_F \\
&\leq ||\mathbf{X}|| \cdot ||(\mathbf{T}^T\mathbf{T})^{1/2} - \mathbf{I}||_F + ||\mathbf{R}|| \cdot ||\mathbf{Y}|| \cdot ||(\mathbf{Y}^T\mathbf{Y})^{-1}||_F
\end{aligned}$$

Now, $||(\mathbf{T}^T\mathbf{T})^{1/2} - \mathbf{I}||_F \leq ||\mathbf{T}^T\mathbf{T} - \mathbf{I}||_F$. Indeed,

$$\begin{aligned}
||(\mathbf{T}^T\mathbf{T})^{1/2} - \mathbf{I}||_F^2 &= \sum_{i=1}^{d}(\lambda_i(\mathbf{T}^T\mathbf{T})^{1/2} - 1)^2 \leq \sum_{i=1}^{d}(\lambda_i(\mathbf{T}^T\mathbf{T}) - 1)^2 \\
&= ||\mathbf{T}^T\mathbf{T} - \mathbf{I}||_F
\end{aligned}$$

We thus have

$$||\mathbf{XW} - \mathbf{Y}|| \leq (||\mathbf{X}|| + ||\mathbf{Y}||)\frac{||\mathbf{R}||\sqrt{d}}{\delta}$$

and **Lemma 1** follows. $\square$

A.2 Procrustean Similarity Transformations

(Borg et al., 2005) Let $\mathbf{X}$ be the target configuration and $\mathbf{Y}$ be the corresponding testee. Assume that $\mathbf{X}$ and $\mathbf{Y}$ are both of order $n \times m$. Let $\mathbf{T}$ be an orthogonal matrix of order $m \times m$. Let δ be a scalar. Let $\mathbf{t}$ be a translation vector. The Procrustes Procedures aim to find $s\mathbf{YT} + \mathbf{1t}^T$ which is close to $\mathbf{X}$. The steps to compute the Procrustean similarity transformations which minimize $||\mathbf{X} - (s\mathbf{YT} + \mathbf{1t}^T)||_F^2$ are:

(1) Compute $\mathbf{C} = \mathbf{X}^T\mathbf{JY}$

(2) Compute the SVD of $\mathbf{C}$; that is, $\mathbf{C} = \mathbf{P\Phi Q}^T$

(3) The optimal rotation matrix is $\mathbf{T} = \mathbf{QP}^T$

(4) The optimal dilation factor is $s = tr(\mathbf{X}^T\mathbf{JYT})/tr(\mathbf{Y}^T\mathbf{JY})$

(5) The optimal translation vector is $\mathbf{t} = n^{-1}(\mathbf{X} - s\mathbf{YT})^T\mathbf{1}$

References

Tang, M., D. L. Sussman, and C. Priebe (2013). "Universally Consistent Vertex Classification For Latent Positions Graphs". In: *arXiv:1212.1182v3*.

Borg, Ingwer and Patrick J. F. Groenen (2005). *Modern Multidimensional Scaling: Theory and Applications*. Springer New York, pp. 429–436.

Jipeng Zhang

108 W. 39th Street | jzhan211@jhu.edu
Baltimore, MD, 21210 | (315)849-6616

EDUCATION

Johns Hopkins University, Whiting School of Engineering — **December 2019**
Master of Science in Engineering. Applied Mathematics and Statistics
Syracuse University, College of Arts and Sciences — **August 2018**
Bachelor of Science. Applied Mathematics

RESEARCH INTERESTS

Network analysis, statistical learning, and imaging processing.

RESEARCH EXPERIENCE

Segmentation of Magnetic Resonance Imaging in Real Time
Baltimore, MD — **5/2019 – Present**

- Built 2D U-Net and 3D U-Net to delineate the outline of brain from MRIs, and to distinguish injected areas from non-injected areas of the brain
- Tune hyperparameters; compare performance of 2D U-Net and 3D U-Net
- Improve segmentation performance on brain margins by combining Generative Adversarial Networks (GANs) with U-Net

Predictive Models for Lung Cancer
Baltimore, MD — **9/2019 – 12/2019**

- Used classification methods including XGBoost, Microsoft Interpret Model, and convolutional neural networks to predict lung cancer from National Lung Screening Trail (NLST) data
- Explained causalities between variables using process mining algorithms Alpha Miner and Heuristic Miner

Spectral Inference for Large Stochastic Blockmodels with Nodal Covariates
Baltimore, MD — **12/2018 – 4/2019**

- Developed algorithm to recover latent positions and parameters of stochastic blockmodels with covariates
- Created R package in GitHub to implement algorithm.

Improving the Transformer Based TSP Solver by Transfer Learning
Baltimore, MD — **2/2019 - 5/2019**

- Built attention-based neural network to solve Travelling Salesman Problem (TSP); trained network through reinforcement learning with baseline
- Boosted learning process by introducing transfer learning into training mechanism

Deep Dive into U-Net: Segmenting MRI
Baltimore, MD — **9/2018 - 12/2018**

- Employed U-Net on Python to segment MRIs
- Analyzed performance of U-Net by altering activation function, normalization, and dropout frequency.

WORK EXPERIENCE

Research Assistant — **10/2019 – 12/2019**
Johns Hopkins University, Bloomberg School of Public Health
Teaching Assistant — **8/2019 – 12/2019**
Johns Hopkins University, Department of Applied Mathematics and Statistics
Research Assistant — **1/2019 - 4/2019**
Johns Hopkins University, Carey Business School